In vivo retinal imaging for fixational eye motion detection using a high-speed digital micromirror device (DMD)-based ophthalmoscope

KARI V. VIENOLA,[1,*] MATHI DAMODARAN,[1] BOY BRAAF,[1] KOENRAAD A. VERMEER,[2] AND JOHANNES F. DE BOER[1]

[1]*LaserLaB, Department of Physics and Astronomy, Vrije Universiteit Amsterdam, De Boelelaan 1081, HV Amsterdam, The Netherlands*
[2]*Rotterdam Ophthalmic Institute, Schiedamse Vest 160D, 3011 BH Rotterdam, The Netherlands*
*k.vienola@vu.nl

Abstract: Retinal motion detection with an accuracy of 0.77 arcmin corresponding to 3.7 µm on the retina is demonstrated with a novel digital micromirror device based ophthalmoscope. By generating a confocal image as a reference, eye motion could be measured from consecutively measured subsampled frames. The subsampled frames provide 7.7 millisecond snapshots of the retina without motion artifacts between the image points of the subsampled frame, distributed over the full field of view. An ophthalmoscope pattern projection speed of 130 Hz enabled a motion detection bandwidth of 65 Hz. A model eye with a scanning mirror was built to test the performance of the motion detection algorithm. Furthermore, an *in vivo* motion trace was obtained from a healthy volunteer. The obtained eye motion trace clearly shows the three main types of fixational eye movements. Lastly, the obtained eye motion trace was used to correct for the eye motion in consecutively obtained subsampled frames to produce an averaged confocal image correct for motion artefacts.

References and links

1. S. Martinez-Conde, S. L. Macknik, and D. H. Hubel, "The role of fixational eye movements in visual perception," Nat. Rev. Neurosci. **5**(3), 229–240 (2004).
2. L. A. Riggs and F. Ratliff, "The Effects of Counteracting the Normal Movements of the Eye," J. Opt. Soc. Am. **42**, 872–873 (1952).
3. R. W. Ditchburn and B. L. Ginsborg, "Vision with a Stabilized Retinal Image," Nature **170**(4314), 36–37 (1952).
4. L. A. Riggs, F. Ratliff, J. C. Cornsweet, and T. N. Cornsweet, "The Disappearance of Steadily Fixated Visual Test Objects," J. Opt. Soc. Am. **43**(6), 495–501 (1953).
5. R. W. Ditchburn, D. H. Fender, and S. Mayne, "Vision with controlled movements of the retinal image," J. Physiol. **145**(1), 98–107 (1959).
6. A. E. Drysdale, "The visibility of retinal blood vessels," Vision Res. **15**(7), 813–818 (1975).
7. J. Jurin, "An essay on distinct and indistinct vision," in *A Compleat System of Opticks, viz. A Popular, A Mathematical, a Mechanical, and a Philosophical Treatise*, R. Smith, ed. (1738), pp. 115–171.
8. E. B. Huey, "Preliminary Experiments in the Physiology and Psychology of Reading," Am. J. Psychol. **9**(4), 575–586 (1898).
9. R. W. Ditchburn and B. L. Ginsborg, "Involuntary eye movements during fixation," J. Physiol. **119**(1), 1–17 (1953).
10. D. A. Robinson, "A method of measuring eye movement using a scleral search coil in a magnetic field," IEEE Trans. Biomed. Eng. **10**, 137–145 (1963).
11. T. N. Cornsweet and H. D. Crane, "Accurate two-dimensional eye tracker using first and fourth Purkinje images," J. Opt. Soc. Am. **63**(8), 921–928 (1973).
12. H. D. Crane and C. M. Steele, "Generation-V dual-Purkinje-image eyetracker," Appl. Opt. **24**(4), 527–537 (1985).
13. M. Barbosa and A. C. James, "Joint iris boundary detection and fit: a real-time method for accurate pupil tracking," Biomed. Opt. Express **5**(8), 2458–2470 (2014).

14. A. B. Roig, M. Morales, J. Espinosa, J. Perez, D. Mas, and C. Illueca, "Pupil detection and tracking for analysis of fixational eye micromovements," Optik **123**(1), 11–15 (2012).

15. T. N. Cornsweet, "New Technique for the Measurement of Small Eye Movements," J. Opt. Soc. Am. **48**(11), 808–811 (1958).

16. R. H. Webb and G. W. Hughes, "Scanning laser ophthalmoscope," IEEE Trans. Biomed. Eng. **28**(7), 488–492 (1981).

17. R. H. Webb, G. W. Hughes, and F. C. Delori, "Confocal scanning laser ophthalmoscope," Appl. Opt. **26**(8), 1492–1499 (1987).

18. J. B. Mulligan, "Recovery of motion parameters from distortions in scanned images," in *NASA Image Registration Workshop (IRW97)*, J. Le Moigne, ed. (NASA Goddard Space Flight Center, Greenbelt, Maryland, USA, 1997).

19. M. Stetter, R. A. Sendtner, and G. T. Timberlake, "A novel method for measuring saccade profiles using the scanning laser ophthalmoscope," Vision Res. **36**(13), 1987–1994 (1996).

20. D. P. Wornson, G. W. Hughes, and R. H. Webb, "Fundus tracking with the scanning laser ophthalmoscope," Appl. Opt. **26**(8), 1500–1504 (1987).

21. R. D. Ferguson, "Servo tracking system utilizing phase-sensitive detection of reflectance variations," U.S. Patent 5,943,115 (August 24, 1999).

22. D. Hammer, R. Ferguson, J. Magill, M. White, A. Elsner, and R. Webb, "Image stabilization for scanning laser ophthalmoscopy," Opt. Express **10**(26), 1542–1549 (2002).

23. Q. Yang, D. W. Arathorn, P. Tiruveedhula, C. R. Vogel, and A. Roorda, "Design of an integrated hardware interface for AOSLO image capture and cone-targeted stimulus delivery," Opt. Express **18**(17), 17841–17858 (2010).

24. C. K. Sheehy, Q. Yang, D. W. Arathorn, P. Tiruveedhula, J. F. de Boer, and A. Roorda, "High-speed, image-based eye tracking with a scanning laser ophthalmoscope," Biomed. Opt. Express **3**(10), 2611–2622 (2012).

25. M. Damodaran, K. V. Vienola, B. Braaf, K. A. Vermeer, and J. F. de Boer, "Digital micromirror device based ophthalmoscope with concentric circle scanning," Biomed. Opt. Express **8**(5), 2766–2780 (2017).

26. K. V. Vienola, M. Damodaran, B. Braaf, K. A. Vermeer, and J. F. de Boer, "Parallel line scanning ophthalmoscope for retinal imaging," Opt. Lett. **40**(22), 5335–5338 (2015).

27. L. J. Hornbeck, "Digital Light Processing for high-brightness high-resolution applications," Proc. SPIE **3013**, 27–40 (1997).

28. K. V. Vienola, B. Braaf, C. K. Sheehy, Q. Yang, P. Tiruveedhula, D. W. Arathorn, J. F. de Boer, and A. Roorda, "Real-time eye motion compensation for OCT imaging with tracking SLO," Biomed. Opt. Express **3**(11), 2950–2963 (2012).

29. E. DeHoog and J. Schwiegerling, "Optimal parameters for retinal illumination and imaging in fundus cameras," Appl. Opt. **47**(36), 6769–6777 (2008).

30. R. Heintzmann, V. Sarafis, P. Munroe, J. Nailon, Q. S. Hanley, and T. M. Jovin, "Resolution enhancement by subtraction of confocal signals taken at different pinhole sizes," Micron **34**(6-7), 293–300 (2003).

31. J. Lewis, "Fast normalized cross-correlation," Vision interface **10**, 120– 123 (1995).

32. International Electrotechnical Commission, Safety of laser products - Part 1: Equipment classification and requirements, (Geneva, Switzerland), IEC-60825–1, (2014).

33. J. Otero-Millan, S. L. Macknik, and S. Martinez-Conde, "Fixational eye movements and binocular vision," Front. Integr. Nuerosci. **8**, 52 (2014).

34. J. M. Findlay, "Frequency analysis of human involuntary eye movement," Kybernetik **8**(6), 207–214 (1971).

35. Q. Yang, J. Zhang, K. Nozato, K. Saito, D. R. Williams, A. Roorda, and E. A. Rossi, "Closed-loop optical stabilization and digital image registration in adaptive optics scanning light ophthalmoscopy," Biomed. Opt. Express **5**(9), 3174–3191 (2014).

36. D. W. Arathorn, Q. Yang, C. R. Vogel, Y. Zhang, P. Tiruveedhula, and A. Roorda, "Retinally stabilized cone-targeted stimulus delivery," Opt. Express **15**(21), 13731–13744 (2007).

37. P. Bedggood and A. Metha, "De-warping of images and improved eye tracking for the scanning laser ophthalmoscope," PLoS One **12**(4), e0174617 (2017).

38. A. E. Salmon, R. F. Cooper, C. S. Langlo, A. Baghaie, A. Dubra, and J. Carroll, "An Automated Reference Frame Selection (ARFS) Algorithm for Cone Imaging with Adaptive Optics Scanning Light Ophthalmoscopy," Transl. Vis. Sci. Technol. **6**(2), 9 (2017).

39. L. J. Van Rijn, J. Van der Steen, and H. Collewijn, "Instability of ocular torsion during fixation: cyclovergence is more stable than cycloversion," Vision Res. **34**(8), 1077–1087 (1994).

40. F. LaRocca, D. Nankivil, S. Farsiu, and J. A. Izatt, "Handheld simultaneous scanning laser ophthalmoscopy and optical coherence tomography system," Biomed. Opt. Express **4**(11), 2307–2321 (2013).

41. D. I. Barnea and H. F. Silverman, "A Class of Algorithms for Fast Digital Image Registration," IEEE Trans. Comput. **C-21**(2), 179–186 (1972).

42. B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," Proceedings of the 7th international joint conference on Artificial intelligence (IJCAI), pp. 674–679 (1981).

43. C. R. Vogel, D. W. Arathorn, A. Roorda, and A. Parker, "Retinal motion estimation in adaptive optics scanning laser ophthalmoscopy," Opt. Express **14**(2), 487–497 (2006).

1. Introduction

To perceive the stationary world, our eyes keep moving constantly. Even when fixating, the eyes continue to move, causing the image of the fixated object to sweep across hundreds of photoreceptors [1–3]. This involuntary motion prevents the light sensitive cones to adapt to a constant stimulus, which would decrease their sensitivity and cause the image on the retina to fade away [4,5]. This is also why blood vessels that are directly on top of our retina are invisible [6]. These movements were first described as "the trembling of the eye" by Jurin in 1738 [7] but today they are more commonly known as fixational eye movements (FEM).

Several different approaches have been developed over the years to observe FEM. The earliest experiments were quite invasive and consisted of *e.g.* a contact lens with a lever attached to it while the eye was anesthetized [8]. Over time the lever was replaced with a small mirror to increase the reflectivity [9] and later magnetic search coils were introduced [10]. The movement of the reflection from the anterior optics can be also used to measure motion non-invasively with demonstrated sampling rates up to 4 kHz [11,12]. However, the most utilized method for motion detection from the anterior part of the eye is pupil tracking [13,14], which typically uses a combination of infrared illumination and an area scan camera to track the pupil. Many of these methods are still actively used for eye movement research.

The first motion detection based on the posterior part of the eye was done by tracking the horizontal movement of a blood vessel in one dimension [15]. The invention of the scanning laser ophthalmoscope (SLO) by Webb *et al.* [16,17] had a significant impact on eye movement research and its benefits for motion detection were appreciated early on. The ability to record retinal images at video-rate allowed detection of eye motion at that same speed [18], followed by motion analysis within a frame [19,20].

Retinal motion detection based on the posterior part of the eye can be split into feature-based and image-based approaches. The feature-based approach was first demonstrated by Ferguson *et al.* [21] and further developed by Physical Sciences Inc. (PSI) [22]. A dithering probing beam and servo tracking system detected the eye motion by monitoring phase changes in the fundus reflectance of a specific feature such as the optic nerve head (ONH) with a reported tracking bandwidth of up to 1 kHz.

In image-based approaches such as SLO, consecutive frames are compared to a template (also known as the reference frame) to obtain the translational shift between consecutive images. This means that the motion detection bandwidth is limited by the frame imaging speed. However, the motion detection bandwidth can be increased by using subsampled frames for the comparison. For example, every consecutive frame can be divided into stripes that are correlated to the template, thus increasing the temporal resolution. Recently, this method was shown in real-time using field-programmable gate arrays (FPGAs) [23,24]. The frame grabber does not wait for the whole frame to be scanned. Instead, the data is acquired in stripes, effectively increasing the motion detection bandwidth to a multiple of the SLO full frame rate and this bandwidth increase has been demonstrated up to 480 Hz [23]. A good comparison showing the differences in tracking methods can be found in a paper by Sheehy *et al.* [24].

In this paper an image-based eye motion detection scheme is presented using our recently developed SLO [25,26] that utilizes digital micromirror device (DMD) technology [27]. The structured illumination allows the acquisition of subsampled snapshots of the retina, which are then compared to a full confocal frame (a reference frame) to estimate the eye motion using cross-correlation analysis. The subsampled snapshots provide simultaneously acquired points distributed over the full field of view (FOV) within an integration time of 7.7 milliseconds, eliminating motion artifacts in the relative position of these points. With this proof-of-principle approach, retinal information is obtained simultaneously within the whole FOV, unlike the stripe-based raster-scan methods where the data is always acquired in a serial manner, and where motion artifacts can be present within the stripe. Moreover, data points within a stripe are clustered closely together, potentially causing the width of the stripe to be a

limiting factor in the detection of the motion perpendicular to the stripe [24,28]. The performance of the motion detection algorithm with subsampled snapshots was characterized using a model eye. Subsequently, eye movement recordings were analyzed from a healthy subject and finally, a motion corrected averaged image is presented, showing the feasibility of motion detection with subsampled frames.

2. Experimental system

In this section the experimental system is first briefly described (section 2.1) followed by an explanation of the motion detection algorithm (section 2.2). Then the system performance analysis using a model eye is presented (section 2.3). Finally, the imaging protocol for *in vivo* measurements is described (section 2.4).

2.1 Image acquisition

To acquire eye motion data a recently developed DMD-based ophthalmoscope (Fig. 1) was used [25]. The system uses an 810 nm wavelength light emitting diode (LED) for illumination and a digital micromirror device (V4100, Vialux GmbH, Germany) for projection of binary patterns. With the DMD, a concentric circle illumination was created consisting of multiple circles. The FOV was imaged by projecting 20 subsampled frames with changing radius of the projected rings at 130 Hz to cover the full FOV and to create a confocal image of the retina. Annular illumination and detection through an aperture [29] allowed the reflections from the central part of the cornea to be rejected and thus reduced the overall background compared to our previous system [26]. To reduce any internal reflections in the system, linear polarizers, a polarizing beam splitter, and a quarter wave plate (QWP) were used. The scattered and reflected light from the retina was detected with a CMOS camera (ace2040-180kmNIR, Basler, Berlin, Germany). For a more detailed description of the system, the reader is referred to [25, 26].
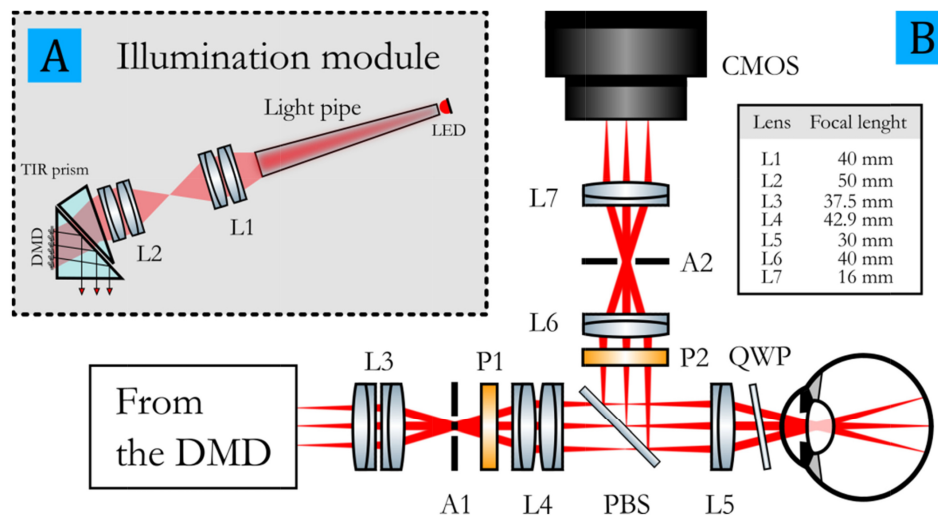

Fig. 1. A schematic of the optical setup. (A) The illumination module consists of LED, light pipe, relay telescope (L1 and L2), total internal reflection (TIR) prism and the DMD. The light pipe homogenizes the LED illumination whereas the TIR prism will direct the light from on mirrors towards the eye. (B) L3-L7: Lenses; P1-P2: Polarizers; A1: Annulus; A2: Circular aperture; PBS: Polarizing beam splitter; QWP: Quarter-wave plate.

To reliably detect all three types of FEM, continuous acquisition was done for several seconds. First, a confocal image was created using the algorithm from Heintzmann *et al.* [30]. After all the patterns from a sequence of 20 were recorded, the maximum and minimum intensity values for each pixel in the sequence were subtracted from each other to reconstruct

a confocal image. The highest intensity values represent the signal in focus, whereas the lowest values are regarded as background signal. This confocal image served as a template (reference frame) which was assumed to contain an insignificant amount of eye motion. If the frame contained eye motion that was obvious when visually inspecting the image, a new reference frame was chosen. Then a cross-correlation analysis was performed to obtain the translational motion using the subsampled snapshots of the retina.

2.2 Motion detection using normalized cross-correlation

The acquired data set was analyzed to detect the (model) eye motion as a function of time. After the acquisition of a full confocal image, the individual subsampled images were compared to the reference frame using a fast normalized cross-correlation (NCC) [31]:

$$\gamma(u,v) = \frac{\sum_{x,y}\left[f(x,y)-\overline{f}_{u,v}\right]\left[t(x-u,y-v)-\overline{t}\right]}{\sqrt{\sum_{x,y}\left[f(x,y)-\overline{f}_{u,v}\right]^2 \sum_{x,y}\left[t(x-u,y-v)-\overline{t}\right]^2}},$$

where $f(x,y)$ is the subsampled frame, $t(x,y)$ the reference frame, and $\overline{f}_{u,v}$ and $\overline{t}$ the respective means. To make the NCC fast, the image correlation is done in Fourier domain after normalization of the images. When the subsampled frame $f(x,y)$ is cross-correlated to the reference frame $t(x,y)$, it will create a two-dimensional (2D) cross-correlation coefficient matrix. The highest value in this matrix indicates the position where the two images correlate the strongest. If no motion occurred between the reference and the subsampled frame this value is in the center of the correlation matrix. However, if there has been motion during the time these two images were acquired, the highest coefficient value is not located in the center of the matrix and this shift from the center gives the amplitude and direction of the motion.

The normalized cross-correlation coefficient is the metric how well the two images correlate with each other. The maximum value for this coefficient is one, which means that the two images are identical (auto-correlation) but possibly shifted. When the correlated image is the complement of the reference image the coefficient gets a value of negative one.

Usually, in the cross-correlation based motion detection algorithms, a threshold is set for this coefficient to indicate when the correlation is getting low. This can happen *e.g.* when the fixation has drifted far away from the original position or when the image contrast has started to degrade for some reason (tear film thinning, head movement). A small correlation coefficient lowers the probability that the highest value found in the matrix is the right peak and therefore the detected shift between the images may not be reliable anymore. However, as we are using subsampled images for correlation, the coefficient will never reach the value of one, but changes based on the fill factor of the DMD. In our case, each subsampled frame consists of several circular line segments that carry sparsely distributed image information over a large retinal area. The subsampled frame areas between the circular line segments do essentially have no information after background signal subtraction, and will therefore have pixel values of (close to) zero. In the cross-correlation these areas therefore do not contribute. In addition, it allows the subsampled frame to move partially outside the borders of the full FOV confocal reference frame without significantly affecting the cross-correlation result. Evidently, the more subsampled frames we combine before correlation, the higher the correlation coefficient will become as each frame will then contain more information but at the same time, it decreases the motion detection bandwidth.

2.3 Model eye

To validate how well the motion detection algorithm works with the system, a model eye was built (Fig. 2(A)) using a single galvanometric mirror (6220H, Cambridge Technology, USA) in the beam path between an uncoated plano-convex lens with a focal length of 15 mm (LA1222, Thorlabs GmbH, Germany) and an artificial retina to generate controlled motion in

the horizontal direction. A function generator (DS345, Stanford Research Systems, USA) generated the driving waveform for controlling the galvo mirror with varying amplitudes and frequencies. To mimic the retinal microstructure a paper business card was used as the retina, which provided random microstructure with minimal specular reflection and a reflectivity comparable to an *in vivo* retina (Fig. 2(B)). The model eye enabled the characterization of the motion detection algorithm by moving the retina with a known frequency and amplitude and subsequently analyzing how well this motion could be extracted from the images. Additionally, it allowed us to measure the effect of parameters such as light intensity, motion frequency and motion amplitude individually, which naturally cannot be done in a human eye as all these effects happen simultaneously.
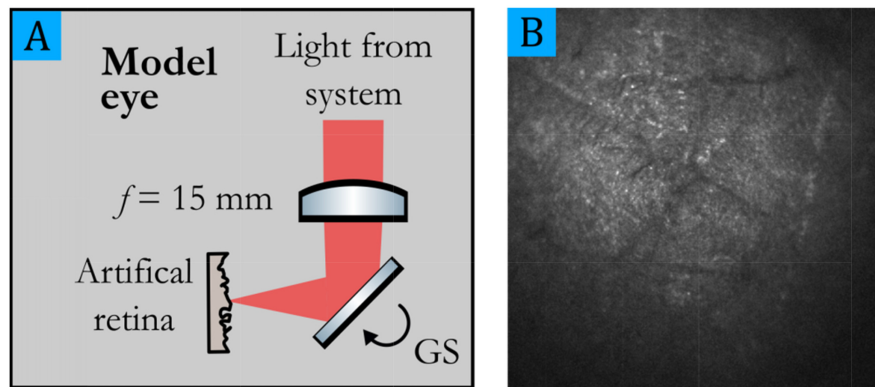

Fig. 2. Model eye with controlled motion. (A) The function generator provided the scanning waveform to the galvo scanner (GS) in the model eye motion separate from the actual imaging system. In the retinal plane, one pixel was about 8 µm and an amplitude of one volt generated a shift of about ± 40 pixels. (B) A confocal image of the surface of the artificial retina.

The driving voltage of the galvanometric mirror was mapped to the image shift in the retinal plane and was measured to be about ± 40 pixels per volt. In the retinal image plane, one camera pixel was measured to be approximately 8.3 µm or 1.9 minutes of arc. All the model eye measurements were performed with the same DMD fill factor of 5% which required a total of 20 different binary patterns to be displayed on the DMD to illuminate the entire FOV. The optical power was kept at 180 µW except in the case of Fig. 4(A) where the cross-correlation peak height was investigated as a function of incident power. Depending on the experiment, the frequency and amplitude of a sinusoidal motion were varied. To obtain an optimal reference frame, the image acquisition always started with a stationary model eye from which the reference was obtained.

2.4 Measurement protocol for in vivo measurements

The use of the experimental setup for *in vivo* measurements in humans was approved by the Institutional Review Board of the VU University Medical Center Amsterdam and adhered to the tenets of the Declaration of Helsinki. Informed consent was obtained from the subject prior to imaging.

To test the motion detection algorithm with the DMD-based ophthalmoscope, a healthy volunteer was imaged. The power in the pupil plane for *in vivo* imaging was measured to be 180-200 µW (using a DMD fill factor of 5%), which is substantially less than the safety limit dictated by the IEC 60825-1 standard for laser safety [32]. The used fill factor resulted in a full image frame rate of 6.5 Hz; however, for the motion detection individual subsampled patterns were projected at 130 Hz making the motion detection bandwidth 65 Hz. For the measurements, no dilation eye drops were used, and the measurements were done in a darkened room with dark-adapted pupils. An individual DMD mirror was magnified to an approximately 7.9 µm spot on the retina [25].

3. System performance (results)

In this section, the obtained results are discussed in detail. First, the algorithm performance was characterized using the model eye with known motion parameters. Then a healthy volunteer was imaged for several seconds and the eye motion was obtained from the data set. A detailed analysis of the ophthalmoscope's imaging performance can be found in a previous publication [25] and is not included in this manuscript.

3.1 Model eye performance

Figure 3 shows two plots that illustrate the retrieved sine wave from the motion detection algorithm in horizontal and vertical directions. The input sine wave had an amplitude of ± 0.5V with a frequency of 10 Hz and the measured NCC peak shift was ± 18.75 pixels with a standard deviation of 0.42 pixels. To achieve subpixel resolution in the cross-correlation, the center of mass of the cross-correlation peak was determined. The horizontal motion is shown in Fig. 3(A) whereas the vertical motion is plotted in Fig. 3(B). To better visualize the retrieved motion trace, only one second of the total motion trace is shown here. As the model eye only generates horizontal motion, the vertical image shift remains close to zero in Fig. 3(B) with a standard deviation of 0.057 pixels (6.5 arcsec) calculated from the whole data set. The sine fit is in good agreement with the retrieved image shift from the algorithm having a fit amplitude of 19.14 pixels and frequency of 10 Hz ($R^2 = 0.9979$). The amplitude of the sine wave corresponds to approximately 33 arcmin whereas typically micro-saccades tend to be less than 0.5° in amplitude, but can be up to 1° or more [1,33]. This demonstrates that our system can measure displacements comparable to eye motion amplitudes that occur in the real human eye. The sinusoidal fit can be subtracted from the experimental data to obtain the accuracy of the detected motion. In Fig. 3(A) this leads to a standard deviation of 0.59 pixels or 1.1 arcmin as the residual motion.
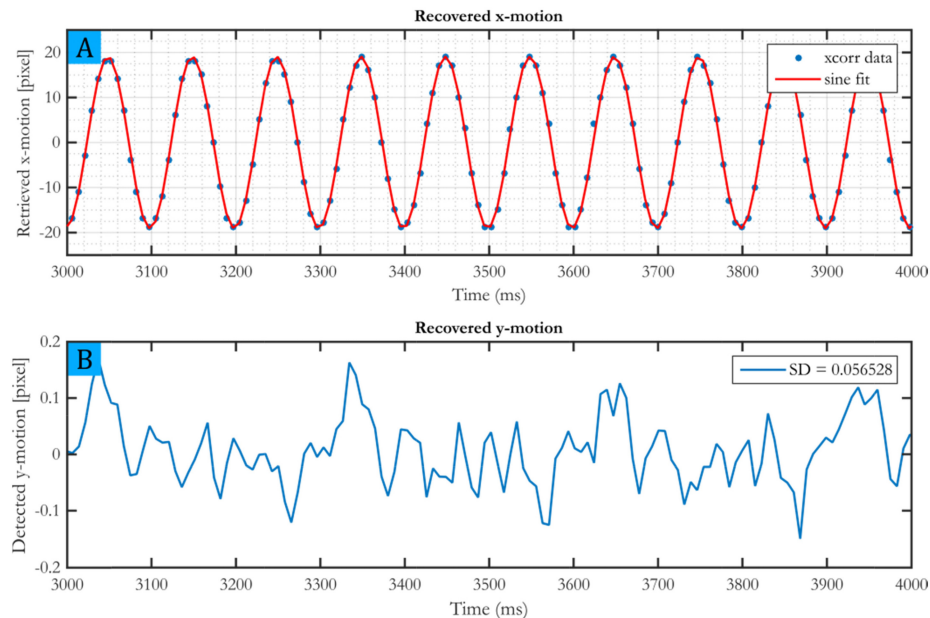

Fig. 3. An example of obtained motion traces using the model eye. (A) The retrieved horizontal motion data is represented in blue dots whereas the sine fit is presented in red line with an extremely good fit having $R^2 = 0.9979$. (B) Because there is no motion in the vertical direction, the motion trace should be zero. However, there is some residual motion coupling from the horizontal channel to vertical channel due to small alignment mismatches of the model eye. The standard deviation of 0.057 pixels (6.5 arcsec) corresponds to about 0.5 μm in the retinal plane. Please note the scaling is different in y-axis between the two plots.

Thinning of the tear film layer is known to reduce the image brightness in a confocal SLO. This thinning of the tear film layer deteriorates the point spread function (PSF), causing more light to be rejected by the confocal detection. This reduction in image brightness caused by the thinning of the tear film layer was mimicked by altering the optical power between 370 and 105 µW sent to the model eye. Figure 4(A) shows the highest correlation coefficient of the matrix as a function of optical power for three different input motion amplitudes, 0 arcmin (dashed line), 6.9 arcmin (red) and 33.0 arcmin (yellow). The cross-correlation coefficients were only calculated at the stationary points of the motion, i.e. the amplitude peaks of the motion trace (such as shown in Fig. 3). The mean value of the coefficient was then plotted as a function of the optical power with the standard deviation as error bars. In all cases, the peak height decreases throughout the curve starting from about 0.35 and reaching approx. 0.25 with the lowest power setting (about 100 µW on the cornea). The standard deviation is caused by the noise in the images (such as shot noise and readout noise). Secondly, the cross-correlation is calculated from varying subframes, with different information content and thirdly, noise is more dominant when the overlap of the subframe with the reference frame decreases.
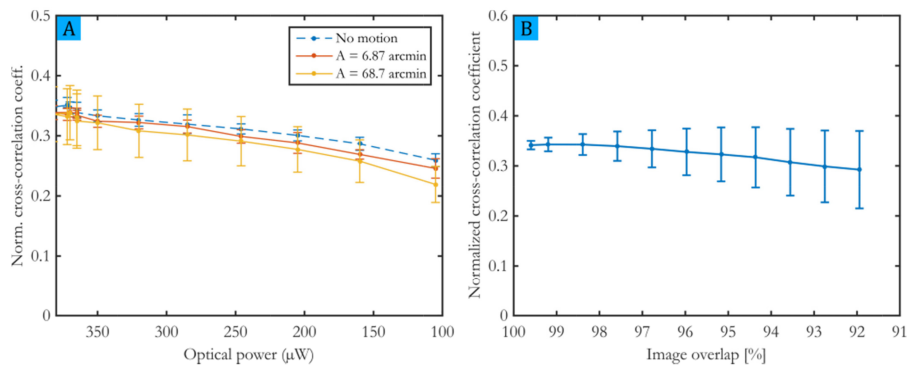

Fig. 4. The normalized cross-correlation coefficient behavior for different conditions. (A) From the obtained motion trace the highest amplitudes were detected (stationary points) and the corresponding correlation coefficient was acquired. The amplitude was kept constant for each curve, but the optical power was varied. The smallest amplitude follows a similar path as the dashed blue line (no motion) whereas the yellow line decreases a bit stronger. (B) As the image overlap decreases (motion amplitude increases) there are only minor changes in the correlation coefficient. However, the standard deviation of the coefficient increases which is seen in the growing error bars when the image overlap is decreased. The 92% overlap corresponds to about 2.28° shift which is much larger than the typical amplitude of a micro-saccade [1].

Lastly, the effect of image overlap in the normalized cross-correlation coefficient values was investigated as there are fewer pixels for correlation when the image shift (motion amplitude) becomes larger. Figure 4(B) shows the normalized cross-correlation coefficient peak value plotted as a function of image overlap (motion amplitude). Again, the stationary amplitude peaks from the trace were located and the corresponding cross-correlation coefficient was obtained. The mean of the coefficients was then plotted as a function of image overlap with a standard deviation showing the error. The maximum input amplitude used was 2 V which corresponds to an image shift of ± 80 pixels (8% shift or 2.28°). The typical micro-saccades rarely go above one degree so within that range the decrease in the coefficient is minimal as can be seen from the plot (overlap from 100% to 96%).

3.2 In vivo eye measurements

An example of an *in vivo* image correlation is shown in Fig. 5. The reference, in this case, is a confocal image of the ONH area. The subsampled frame of the same area is then cross-correlated with the reference frame resulting in a 2D correlation matrix seen on the right side

of Fig. 5. With the 5% fill factor of the DMD, the highest NCC coefficient values reach about 0.3 to 0.35. For the *in vivo* data we did not determine the center of mass of the cross-correlation, but used the peak location, resulting in a motion resolution determined by the image pixel size (corresponding to 7.9 µm on the retina or 1.6 arcmin).

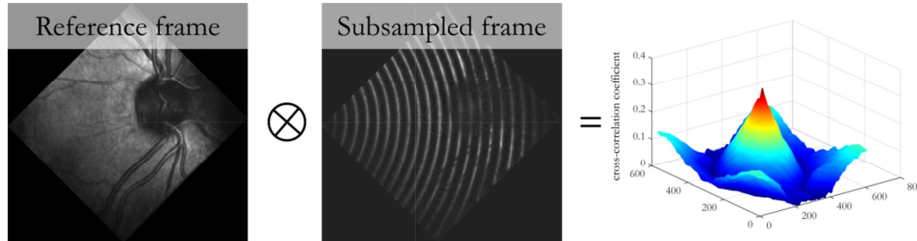Fig. 5. An *in vivo* example of the cross-correlation. First a confocal frame is constructed, which will act as a reference frame. Then the next subsampled frame will be cross-correlated to the reference frame to obtain the shift between these two frames. The peak that occurs in the correlation matrix indicates the offset of the subsampled pattern with respect to the reference. The better the two images match, the higher the peak in the correlation matrix will be.

An *in vivo* eye motion trace is presented in Fig. 6, where the red line indicates motion in the horizontal direction and the blue trace shows motion in the vertical direction. All 3 types of FEM can be distinguished from the trace. Two micro-saccades occurred during the recording at around 1.1 seconds and at 2.9 seconds, having amplitudes of about 17 and 22 arcmin respectively. Slow drift is clearly visible throughout the trace having amplitude values between 5 and 20 arcmin, which is like reported in the literature [1]. And finally, the tremor is visible superimposed on top of the motion trace as high frequency low amplitude jitter.

Since the true eye motion is not known, we estimated an upper bound of the motion detection accuracy as follows. We assumed that the location of each subframe within the reference frame was correctly found within the motion detection accuracy. So, in the absence of motion, each subframe jitters around the true position within the motion detection accuracy. The "true" position of the frame was estimated by averaging over the position of 7 neighboring frames. The upper bound of the motion detection accuracy was determined by calculating the standard deviation of each frame position with respect to the average over 7 neighboring frames. For this measurement the micro-saccades were removed from the motion trace. The acquired values for the standard deviation were 0.55 arcmin for the horizontal trace and 0.53 arcmin for the vertical trace, respectively. Combining the horizontal and vertical standard deviations into a 2D radial standard deviation gives 0.77 arcmin, corresponding to a radial standard deviation of 3.7 µm on the retina. Visualization 1 shows the confocal video generated from the data (left) and the corresponding eye movement is drawn in the plot on the right.
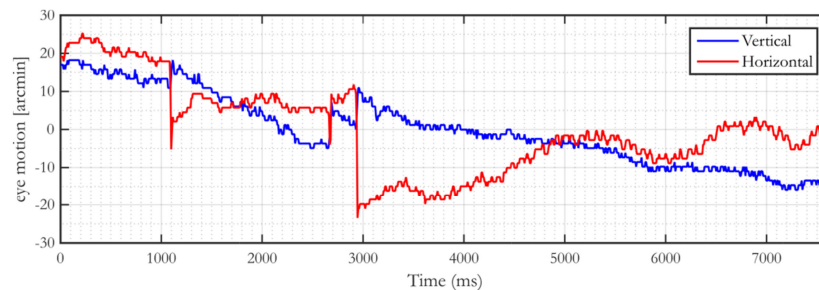

Fig. 6. Extracted eye motion traces from a healthy subject. All three types of eye motion can be distinguished from the traces, namely micro-saccades (large jumps at 1.1 s and 2.9 s), drift (drifting motion along the trace with small amplitude and frequency) and tremor, high frequency motion superimposed on top of the eye motion trace (Visualization 1).

When the eye motion amplitude is plotted as a function of frequency (Fig. 7), it shows a similar behavior to $1/f$. This same behavior has been reported in the literature on many occasions [24,34] and it supports the validity of the eye motion trace. It also shows that the tremor component happens more frequently but has much smaller amplitude than the large amplitude micro-saccades. The motion detection bandwidth allows us to detect motion up to 65 Hz as dictated by the Nyquist theorem.
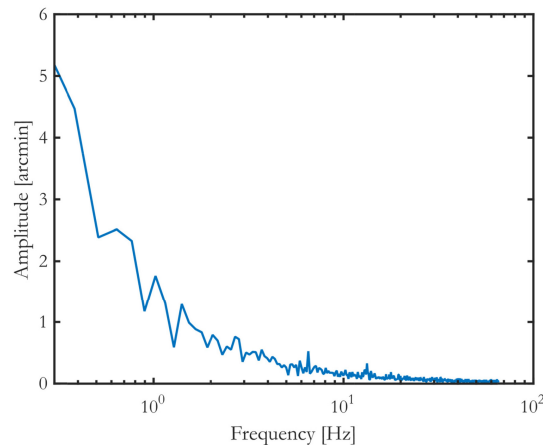

Fig. 7. Eye motion amplitude as a function of frequency. The spectrum follows the well-known $1/f$ curve and ends at 65 Hz which is the current motion detection bandwidth. Low frequency motion such as drift and micro-saccades have larger amplitude than high frequency tremor, which is present up to the detection limit. The two peaks seen at 6.5 Hz and 13 Hz are artefacts that originate from the reference frame (see [28]).

The obtained eye motion trace shows the displacement of each subsampled frame compared to the reference frame. This displacement information can be used to adjust the position of each subsampled frame with respect to the reference frame and generate an averaged confocal image that is corrected for motion. For this specific data set, a total of 1000 subsampled frames were taken. With a DMD fill factor of 0.05, this results in 50 confocal full images (20 patterns per confocal image) over 7.6 seconds with most of them affected by eye motion.
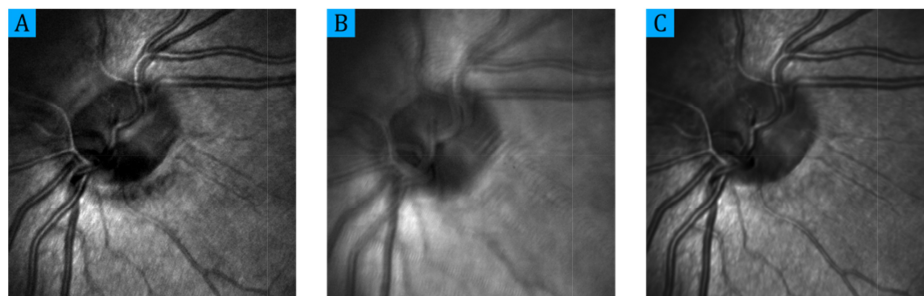

Fig. 8. Averaging multiple confocal images without and with motion correction. To generate the images, 1000 subsampled frames were taken over 7.6 seconds. As the fill factor of the DMD was 0.05, it took 20 patterns to scan the entire FOV. This then resulted in 50 full confocal images that were averaged. (A) A single confocal image for comparison. (B) When the subsamples images are not corrected for motion, the resulting averaged image is blurry. (C) When each subsampled frame is corrected for the eye motion, the averaged image has high quality, showing good contrast and lots of features typical of the area around the ONH.

Figure 8 shows three confocal images. In Fig. 8(A), a single confocal image is shown as a reference. In Fig. 8(B) the patterns are not corrected according to the eye motion trace before

applying the Heintzmann algorithm and the resulting confocal image is blurry and distorted by the eye motion. In Fig. 8(C), the eye motion during the measurement has been corrected before applying the algorithm. The resulting high-quality image shows good contrast and many anatomical features such as the larger blood vessels that are now sharp and the smaller vessels originating from ONH that are visible.

4. Discussion

To summarize the results, the DMD-based ophthalmoscope was able to detect *in vivo* eye motion at 130 Hz, reaching a motion detection bandwidth of 65 Hz. Although the tremor frequency range can be larger than the system's motion detection bandwidth [1], the accumulative image shifts due to multiple tremor motions can be still detected and corrected. In comparison to other systems, our system does not yet reach the same eye motion detection bandwidth [22,23,35] but we were able to extract motion using subsampled frames within the whole FOV, i.e., using information isotropic in both horizontal and vertical direction. This means that the motion detection works equally well for horizontal motion as for vertical motion, which has typically been a concern in the stripe-based method [36]. Moreover, each subsampled frame is a snapshot and has no motion distortion artifacts between the points in the subsampled frame, although it might suffer from some blurring due to rapid motion during the subsampled frame's integration time. Lastly our system operates below the performance of TSLO in accuracy and this is partly due to our system having a larger FOV making our sampling density sparser compared to the TSLO.

The model eye measurements showed that the normalized cross-correlation method works well with the subsampled images, keeping the correlation peak above the noise floor for reliable motion detection and that the change in optical power or image overlap has only little effect on the motion detection in the motion regime that is relevant. The detection of incorrect peaks can be minimized by setting up a threshold value for the normalized cross-correlation to make the tracking more reliable. The accuracy of the motion detection is slightly less than with the *in vivo* measurements. We hypothesize that this is related to the contrast in the confocal image. With better contrast, the cross-correlation is sharper peaked around the correct shift. The *in vivo* eye motion traces were recorded centered on the optic nerve head, a feature and contrast rich area. Whereas stripes in the stripe-based motion detection method could be in retinal areas with low contrast and few features, the snapshot-based subsampled frames provide information over the full FOV, better sampling the feature and contrast rich areas.

Several improvements are still possible. Currently we are using 200 µW of power in the cornea and this could be increased by a factor of 3.5 to provide more signal from the retina as our SNR calculations [25,26] demonstrate, based on a single stationary spot focused on the retina. If the sub frame illumination is evaluated as a semi-extended source compared to our conservative collimated beam approach, maximum permissible optical powers could be significantly increased. Especially the use of annulus in the imaging path reduces the optical power in our system, not to mention the low fill factor of the DMD. With more powerful illumination sources, the circular patterns can be projected even faster, and the frame integration time can be lowered to about a millisecond, giving a full confocal imaging at 50 Hz and increasing the motion detection bandwidth to 500 Hz.

Furthermore, the reference frame might not be always artifact free. To resolve this, recently published methods from Bedggood and Metha [37] or Salmon *et al.* [38] could be implemented to obtain a better reference frame estimate. Lastly, the cross-correlation detects any translation in x- and y-plane. Torsional eye motion in a fixating eye has a standard deviation of less than 0.25 degrees [39] and so has a relatively small contribution on the time scales measured here. Efficient computation algorithms could be implemented to compensate also for small rotations with x- and y-movement such as the method from LaRocca *et al.* [40].

For the future, other mapping approaches could be also tested such as the sequential similarity detection algorithm (SSDA) [41] or gradient descent search [42] to further validate if the cross-correlation truly is the best method for image-based motion detection in retinal imaging. Previously Vogel *at al*. demonstrated motion detection with a map-seeking circuit (MSC) with promising results [43] and this should also be considered.

5. Conclusions

In conclusion, we have demonstrated the feasibility of eye motion detection in our DMD-based ophthalmoscope. The motion detection bandwidth was 65 Hz which was enough to detect most of the eye motion in the data set. The motion detection accuracy was better than 0.77 arcmin *in vivo* corresponding to 3.7 µm on the retina. Our system with motion detection capabilities can be used for studying the eye motion and at the same time it will help the averaging of images for improved SNR as the images can be registered together with higher accuracy based on the motion information. Our method provides an alternative to the image-based motion detection done with SLO and in the future, it could be incorporated to other imaging modalities such as optical coherence tomography (angiography) and provide tracking capabilities as an auxiliary device.

Funding

Disclosures

JFdB: Heidelberg Engineering, GmBH (F, P, R), KAV: (P).